

---

# Выделение элементов реакции на электроэнцефалограмме

---

A Preprint

Никишкина Евгения Геннадьевна  
nikishkina.jane@gmail.com

Московский государственный университет имени М. В. Ломоносова  
Научный руководитель: Майсурадзе Арчил Ивериевич

## Abstract

Изменения в эмоциональном состоянии человека могут влиять на сигналы электроэнцефалограммы (ЭЭГ). Распознавание и анализ эмоций с использованием биологических сигналов мозга является сложной задачей, требующей точной обработки и выделения признаков для определения наличия эмоции у конкретного человека. Целью данной работы было создание эффективного алгоритма для распознавания эмоций человека на основе анализа рядов ЭЭГ. Были проанализированы различные подходы, в результате чего был выбран оптимальный метод, который затем был реализован в собственной программе. Также были проведены эксперименты на основе подобранных рядов ЭЭГ, содержащих данные об эмоциональных состояниях. Для детектирования эмоций в рядах ЭЭГ был использован новый подход, основанный на анализе аномалий. Результаты экспериментов продемонстрировали, что разработанный алгоритм позволяет получать интерпретируемые результаты по определению наличия эмоций у человека.

Keywords Электронцефалография (ЭЭГ) · Временные ряды · Детектирование аномалий · Распознавание эмоций · SEED dataset · Глубинное обучение

## 1 Введение

Эмоции представляют собой проявления не только физиологических состояний, связанных с разнообразными чувствами, мыслями и поведением людей, но и психологических откликов на множество внешних стимулов. Умение точно распознавать эмоции играет ключевую роль во множестве сфер: взаимодействие человека и компьютера (HCI) [Satyajit Nayak, 2021], здравоохранение [Valentina Colonnello, 2018], интернет-образование [Xiang Feng and Ma, 2020], мониторинг безопасности [Erkang Fu, 2021], психологический анализ [Wang, 2020], а также индустрия развлечений [Regan L. Mandryk, 2006]. Например, распознавание эмоций может оказаться весьма полезным для диагностики таких психических заболеваний, как депрессия и шизофрения [Bruce I. Turetsky, 2007]. Также при взаимодействии человека с компьютером извлеченные эмоции могут использоваться как своего рода обратная связь для предоставления лучшего контента, улучшающего пользовательский опыт в электронном обучении, компьютерных играх и поиске информации ([Satyajit Nayak, 2021]).

В последние годы существующие модели распознавания эмоций были разделены на две категории: методы, основанные на физиологических сигналах, и методы, основанные на нефизиологических сигналах [L. Shu and Yang., 2018]. По сравнению с нефизиологическими сигналами, физиологические сигналы не восприимчивы к субъективным факторам, которые могут достоверно отображать эмоциональные состояния человека. Таким образом, распознавание эмоций, основанное на физиологических сигналах, имеет большие преимущества в надежности и практичности [Yujian Cai and Li, 2023]. Временный ряд представляет собой собранный в разные моменты времени статистический материал о значении каких-либо параметров исследуемого процесса [Stephen Haben, 2023]. Каждая единица статистического материала называется измерением или отсчетом, причем для каждого отсчета должно быть указано время или порядковый номер измерения. Важно отметить, что при анализе временного ряда учитывает-

ся взаимосвязь измерений со временем, что отличает его от простой выборки данных, где учитывается только статистическое разнообразие и характеристики выборки. В области исследований, основанных на физиологических сигналах, ЭЭГ - это спонтанный, несубъективный физиологический сигнал, который может объективно отражать эмоциональные состояния человека [Mohamed Namada, 2018]. В данной работе объектом исследования являются реальные временные ряды, а именно, — записи электроэнцефалограммы. Электроэнцефалография (ЭЭГ) является методом исследования электрической активности мозга путем размещения электродов в определенных зонах на поверхности головы. Данные, полученные таким образом, представляют собой многомерный временной ряд (каждый датчик отвечает за одномерный временной ряд), то есть система из нескольких одномерных рядов, где целевые значения зависят не только от предыдущих значений одного временного ряда, но и от значений дополнительных временных рядов [Zheng et al., 2018].

Методы распознавания эмоций на основе ЭЭГ можно разделить на две группы: методы классического машинного обучения и глубокого обучения [L. Shu and Yang., 2018]. В методах распознавания эмоций, основанных на классическом машинном обучении, признаки извлекаются вручную для последующего использования в статистических методах, метрических, модельных методов, таких как, например, метод опорных векторов (SVM) Schölkopf et al. [1999] и Isolation Forest Liu et al. [2008]. Лин и др. L. Shu and Yang. [2018] представили общий процесс традиционных методов машинного обучения для распознавания эмоций по ЭЭГ, включая регистрацию эмоций, получение сигнала, выделение признаков, последующие распознавание или классификация эмоций и т.д. Однако, с такими подходами возникает проблема: из-за нелинейности и высокой размерности сигналов ЭЭГ, методы, эффективные на линейных и низкоразмерных данных, не обеспечивают высокое качество решения сложных задач. Также данные подходы имеет существенный недостаток, заключающийся в том, что процесс извлечения признаков обычно затруднителен и в значительной степени зависит от экспертов-людей. Трудно получить все подразумеваемые признаки путем извлечения вручную, и используемые формулы часто очень сложны. Кроме того, сигналы ЭЭГ чувствительны к шумам, таким как электромиографические артефакты, которые создают серьезные помехи. Ввиду вышеуказанных трудностей для решения этих проблем используются некоторые методы глубокого обучения. Альхагри и др. [Salma Alhagry, 2017] предложили LSTM для изучения особенностей по сигналам ЭЭГ, затем в качестве классификатора используется полносвязный слой. Для оценки представленного метода использовался набор данных DEAP, на котором предложенный метод дает среднюю точность 85,65%, 85,45% и 87,99% для классов возбуждения, валентности и симпатии соответственно. Предложенный метод обеспечил высокую среднюю точность по сравнению с традиционными методами. Ширрмейстер и др. [Robin Tibor Schirrmeister, 2017] использовали сверточные нейронные сети (CNN) для декодирования и визуализации ЭЭГ, которые продемонстрировали большой потенциал при применении к сквозному распознаванию эмоций на основе ЭЭГ-сигналов, тем самым подтвердив превосходство глубоких нейронных сетей в задачах классификации ЭЭГ. В последние годы CNN ([Yea-Hoon Kwon and Kim, 2018], [Daksh Maheshwari, 2021]), LSTM ([Salma Alhagry, 2017], [Sharma and Sircar, 2020]), Generative Adversarial Network (GAN) ([Yun Luo and Lu, 2018]) и другие сетевые модели широко использовались при распознавании эмоций по ЭЭГ.

## 2 Постановка задачи

Дан многомерный временной ряд длины  $n$   $X = \{x_1, x_2, x_3, \dots, x_n\}$ , где каждый элемент  $x_t \in \mathbb{R}^m$  -  $m$ -мерный вектор, представляющий собой показания  $m$  датчиков в момент времени  $t$ . Пусть дана обучающая выборка, которая представляет собой многомерные временные ряды для  $\tilde{n}$ , полученные по участникам эксперимента, находящимся в безэмоциональном состоянии. Необходимо для исходного временного ряда предсказать  $Y = \{y_1, y_2, y_3, \dots, y_n\}$ , где каждый элемент  $y_t \in \{0, 1\}^m$  -  $m$ -мерный вектор, определяющий для каждого из  $m$  датчиков аномальность показания в момент времени  $t$ . Будем считать, что момент времени  $t$  аномальный, если показание хотя бы одного из  $m$  датчиков является аномальным. Таким образом, получим вектор  $\tilde{Y} = \{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \dots, \tilde{y}_n\}$ , где элемент  $\tilde{y}_t \in \{0, 1\}$  (1 отвечает за аномальность).

Для тестовых данных для каждой временной точки на основе порога (алгоритм POT) можно получить ответ алгоритма — является ли данная точка аномальной или нет. Разметка для сигналов ЭЭГ отсутствует, поэтому для оценки качества используется косвенная метрика: на основе меток аномальности, полученных алгоритмом, для положительных и нейтральных видео в тестовых данных можно посчитать процент аномальных точек в нем; данную величину можно рассматривать в качестве оценки принадлежности объекта (то есть временного ряда ЭЭГ для конкретного видео) к классу нейтральных данных; на основе этих оценок можно воспользоваться метриками для задачи бинарной классификации (в данной работе была выбрана метрика AUC-ROC).

### 3 Эксперименты

#### 3.1 Данные

Эксперименты проводились на датасете SEED Zheng et al. [2018]. При создании данного датасета были выбраны пятнадцать фрагментов китайских фильмов, характеризующие положительные, нейтральные и отрицательные эмоции в качестве стимулов, использованных в экспериментах. Критерии отбора видеоклипов следующие:

- продолжительность всего эксперимента не должна быть слишком большой, чтобы не вызвать у испытуемого усталость;
- видеоролики должны быть понятны без объяснений;
- видеоролики должны вызывать единственную желаемую целевую эмоцию.

Продолжительность каждого фрагмента фильма составляет примерно 4 минуты. Каждый видеоклип отредактирован так, чтобы создать связную эмоцию, вызывающую интерес, и максимизировать эмоциональный смысл.

В экспериментах принимали участие пятнадцать китайских испытуемых (7 мужчин и 8 женщин). Для каждого эксперимента предусмотрено в общей сложности 15 испытаний. Перед каждым клипом была 5-секундная подсказка о начале, 45 секунд для самостоятельной оценки и 15 секунд для отдыха после каждого клипа в течение одного сеанса. Порядок представления организован таким образом, что два фрагмента фильма, отвечающих за одну и ту же эмоцию, не показываются последовательно. Для получения обратной связи участникам было предложено сообщить о своих эмоциональных реакциях на каждый фрагмент фильма, заполнив анкету сразу после просмотра каждого ролика. Сигналы ЭЭГ регистрировались с помощью 62-канальной системы ESI NeuroScan. Сцена эксперимента и соответствующее расположение электродов для ЭЭГ показаны ниже.

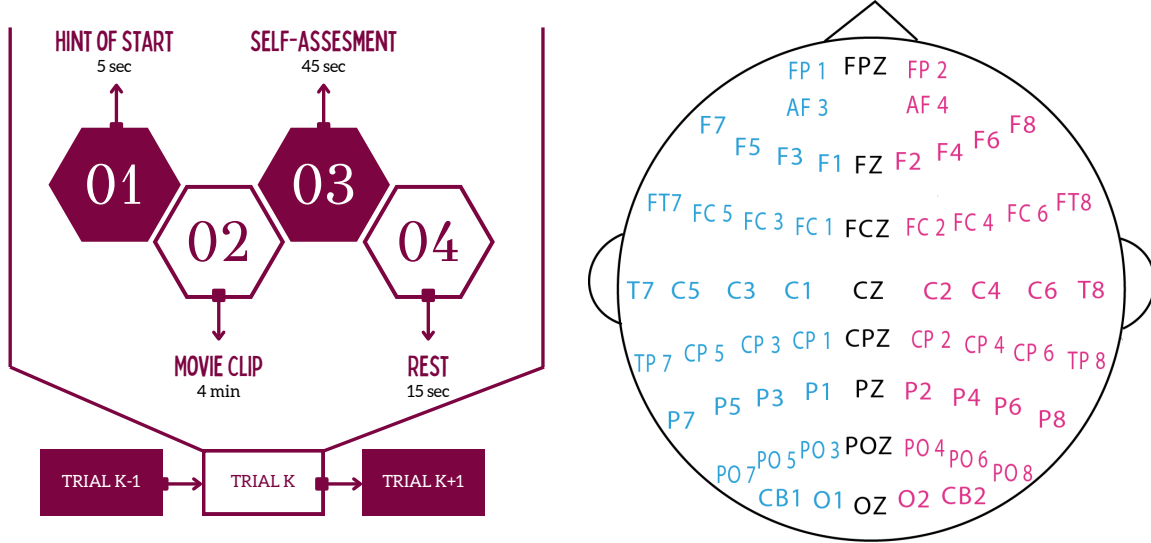


Рис. 1: Сцена эксперимента и расположение электродов для ЭЭГ

Данные представляют собой файлы, содержащие уменьшенные, предварительно обработанные и сегментированные показания ЭЭГ в MATLAB (файл .mat). Частота была уменьшена до 200 Гц., также был применен полосовой фильтр (BPF) с частотой от 0 до 75 Гц, после чего были извлечены сегменты ЭЭГ, соответствующие продолжительности каждого фрагмента фильма. Каждый испытуемый принимал участие в эксперименте три раза с интервалом примерно в одну неделю. Для одного эксперимента было предусмотрено 15 испытаний (1 испытание — просмотр одного видеоролика). Таким образом, для каждого участника эксперимента было получено 45 файлов с расширением .mat, по одному на каждое испытание.

### 3.2 Вычислительные эксперименты

Для того чтобы смоделировать локальную зависимость данных в точке  $x_t$  в момент времени  $t$  необходимо взять окно длины  $K$ :

$$W_t = \{x_{t-K+1}, \dots, x_t\}.$$

Таким образом, весь временной ряд  $T$  преобразуется в последовательность скользящих окон  $W = \{W_1, \dots, W_T\}$ . Для случая  $t < K$  первый элемент временного ряда дублируется такое число раз, чтобы длина окна стала равна  $K$ . Помимо этого с целью восстановления долгосрочных взаимосвязей необходимо также рассматривать временной отрезок до текущей временной метки  $t$  ряда  $T$  (обозначим его  $C$ ).

### 3.3 Базовый эксперимент

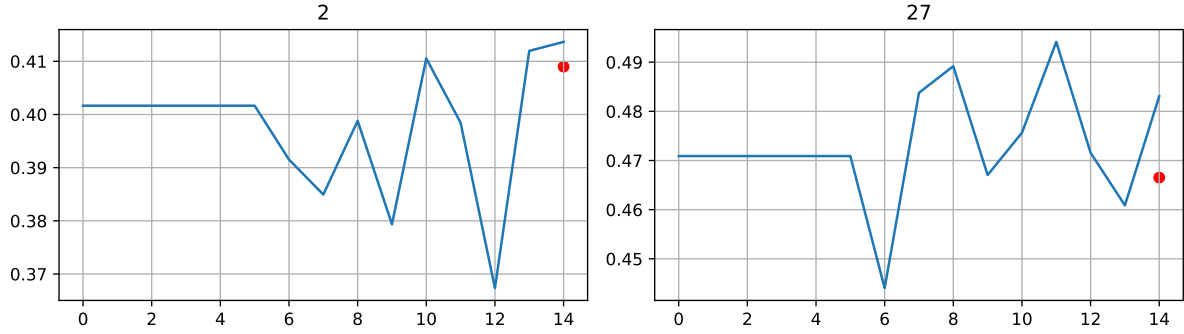


Рис. 2: Полученные результаты предсказания для 2-ого и 27-ого датчиков

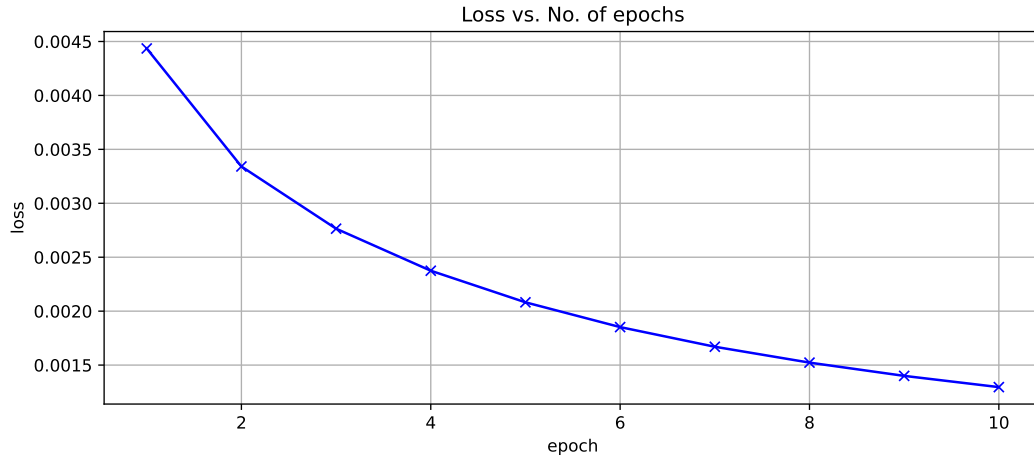


Рис. 3: Зависимость ошибки от номера эпохи на обучении

В качестве базовой модели была взята простая нейросетевая архитектура, состоящая из двух полносвязных слоев и функции активации ReLU.

Обучение состоит в предсказании следующей точки временного ряда, на основе фиксированного количества  $K - 1$  предыдущих точек данного временного ряда. Модель обучается на данных, отвечающих нейтральным эмоциям на задачу минимизации среднеквадратичной ошибки MSE. На рисунке 2 представлено предсказание  $K$ -ой метки временного ряда на основе предыдущих для 2-ого и 27-ого датчика.

На основе такого подхода при получении результатов для вновь поступившего временного ряда, отвечающего яркой эмоции, то есть положительной или отрицательной, выдвигается гипотеза, что поскольку

обучение происходило на данных, отвечающих нейтральным эмоциям, возможен подбор порога на основе которого мы будем определять наличие эмоции у человека. Модель демонстрирует не самые высокие показатели качества 3, поэтому стоит обратить внимание на альтернативную модель, способную обеспечивать более высокое качество работы.

### 3.4 Основной эксперимент

Для детектирования аномалий будем использовать модель TranAD Tuli et al. [2022]. Принцип работы модели схож с принципом работы автокодировщика для детектирования аномалий. Первоначально модель будет переводить переданное на вход окно в пространство меньшей размерности с помощью кодировщика, а затем с помощью декодирующего восстанавливать это окно. Преимущество данного подхода заключается в использовании временного отрезка, который за счет учета долгосрочных тенденций позволяет более точно восстанавливать окно. Также к преимуществам данной модели можно отнести специальную соревновательную процедуру обучения, благодаря которой зашумленные фрагменты предсказываются лучше и не идентифицируются как аномальные.

#### 3.4.1 Модель трансформер

В качестве основных блоков в модели используются кодировщик и декодирующий трансформера Vaswani et al. [2017]. Архитектура модели TranAD представлена на рисунке выше. Для получения представления окна, которое подается на вход декодирующего для его последующего восстановления, используется кодировщик (Window Encoder).

Пусть размерность временных рядов равна  $m$ . Необходимо определить стандартные для трансформерных моделей операции. Скалярным вниманием будем называть следующую величину, полученную в результате операций над матрицами  $Q$  (запросы),  $K$  (ключи) и  $V$  (значения):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{m}}\right)V$$

В модели TranAD используется многоголовое внимание:

$$MultiHeadAtt(Q, K, V) = Concat(H_1, \dots, H_h)$$

где  $H_i = Attention(Q_i, K_i, V_i)$ .

Здесь  $Q_i, K_i, V_i$  — полносвязные слои для каждой из голов трансформера. Дополнительно, чтобы учесть то, что события происходят в разные моменты времени, используется позиционное кодирование Vaswani et al. [2017].

Для обучения модели используется соревновательное обучение наподобие того, что используется в модели GAN. Модель TranAD состоит из двух трансформерных кодировщиков и двух декодирующих. На вход подаются окно  $W$ , ряд  $C$  и специальная матрица focus score (подробнее про него в секции 4.2.2)  $F$ , размеры которой совпадают с размерами окна. Матрица  $F$  дополняется нулями, производится операция конкатенации полученной матрицы  $F$  и ряда  $C$ , затем применяется позиционное кодирование. Полученная матрица  $I_1$  подается на вход первого кодировщика:

$$\begin{aligned} I_1^1 &= LayerNorm(I_1 + MultiHeadAtt(I_1, I_1, I_1)), \\ I_1^2 &= LayerNorm(I_1^1 + FeedForward(I_1^1)). \end{aligned}$$

Перед применением window encoder к окну применяется позиционное кодирование и полученная матрица  $I_2$  подается на его вход. При этом здесь используется маскированное «самовнимание» (mask self-attention) для того, чтобы декодирующий не мог "подсматривать" в будущее во время обучения, то есть во временные точки из будущего в контекстном представлении ряда  $C$ . Таким образом, window encoder представляет собой:

$$\begin{aligned} I_2^1 &= Mask(MultiHeadAtt(I_2, I_2, I_2)), \\ I_2^2 &= LayerNorm(I_2 + I_2^1), \\ I_2^3 &= LayerNorm(I_2^2 + MultiHeadAtt(I_1^2, I_1^2, I_2^2)). \end{aligned}$$

То есть контекстные представления всей последовательности  $C$  используются в качестве ключей и значений для перекрестного внимания в window encoder. Добавление представления всей последовательности  $C$  до момента времени  $t$  позволяет модели использовать расширенный контекст. Следующими блоками в модели являются два одинаковых декодирующего, восстанавливающих переданное окно:

$$O_i = Sigmoid(FeedForward(I_2^3)), i \in \{1, 2\}.$$

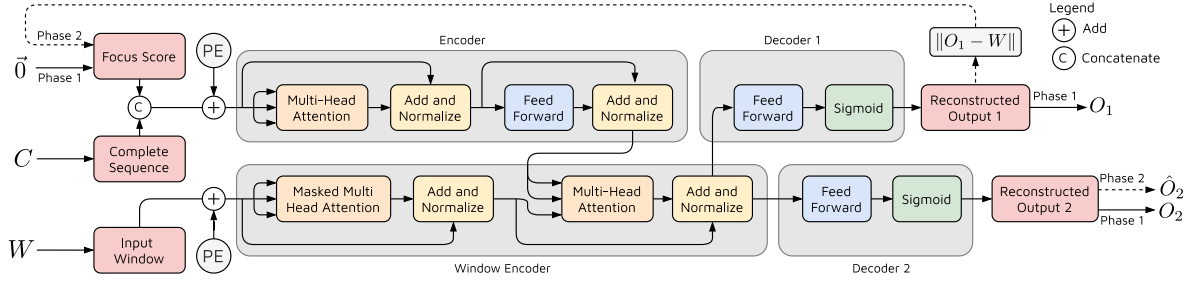


Рис. 4: Модель TranAD

Таким образом, модель TranAD принимает на вход  $C$  и  $W$  и генерирует два выхода  $O_1$  и  $O_2$ .

### 3.4.2 Двухэтапное состязательное обучение

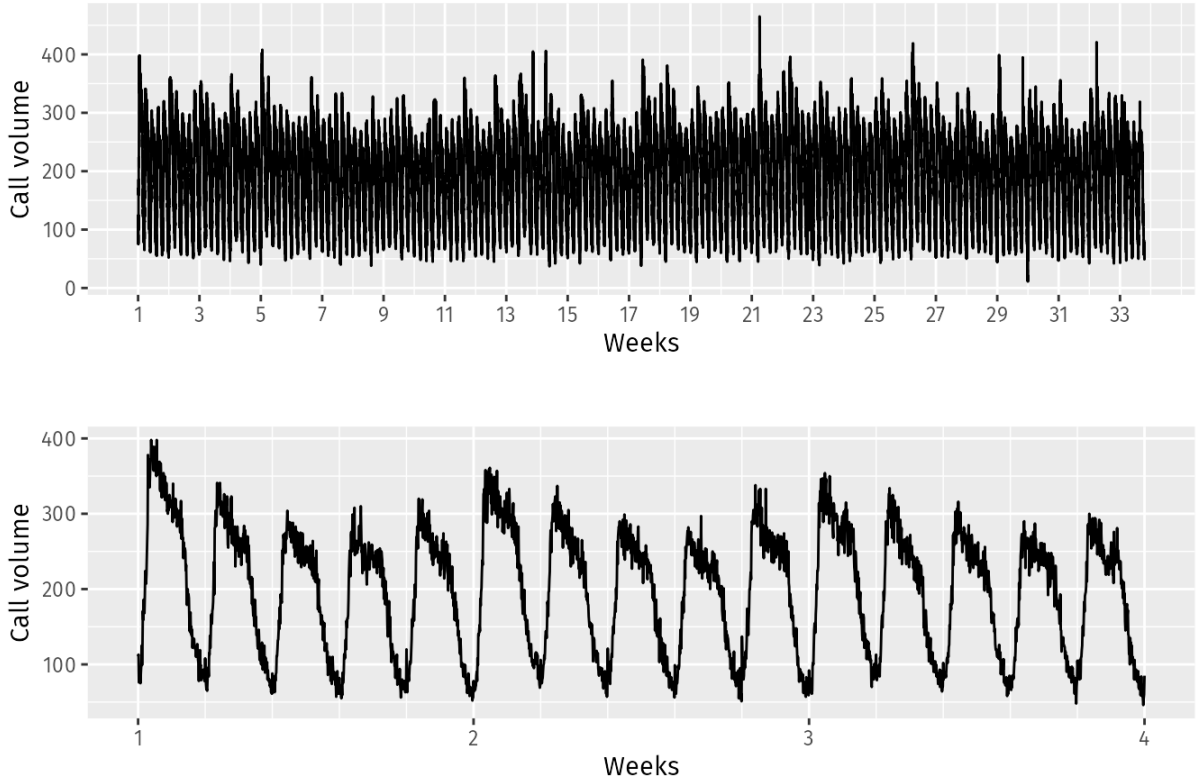


Рис. 5: Пятиминутный объем звонков в крупном североамериканском коммерческом банке

Благодаря учету долгосрочных зависимостей модель способна лучше восстанавливать переданное на вход окно.

Для того чтобы лучше понять, что имеется в виду, рассмотрим следующий пример. На рисунке 5 представлены два графика: на первом графике показано количество поступивших звонков в банки с 5-минутным интервалом в промежутке между 7:00 и 21:05 каждого буднего дня в течение 33 недель, на втором показаны первые три недели этого же временного ряда. Знание всей последовательности до текущего момента  $t$  дает, например, информацию о времени суток, что в свою очередь, потенциально помогает в восстановлении количества звонков в момент времени  $t$ . Заметим, что значения количества

звонков в дневное время суток более шумные, поэтому модели будет сложнее восстанавливать в данный временной промежуток и она может выдать ложноположительный результат. С целью решения данной проблемы и вводится процедура соревновательного обучения.

Процедура соревновательного обучения состоит из двух фаз. Первая фаза представляет собой реконструкцию входа модели. Модель пытается сгенерировать входное окно как можно более точно. Поэлементная разность между сгенерированным первым декодером окном и входным окном и есть focus score, о котором было упомянуто выше. Операция конкатенации focus score и входной последовательности облегчает механизму внимания извлекать локальные зависимости, фокусируясь на подпоследовательностях с высокими значениями focus score. Таким образом, после первой фазы кодировщик преобразует входное окно  $W \in \mathbb{R}^{K \times m}$  (с focus score  $F = [0]_{K \times m}$ ) в сжатое представление  $I_2^3$ , из которого впоследствии получаются выходы  $O_1$  и  $O_2$ .

Вторая фаза представляет собой реконструкцию входа с использованием focus score. Матрица  $F$ , полученная после первой фазы, используется вторым декодером для генерации выхода  $\hat{O}_2$ .

Такое двухстадийное авторегрессионное решение имеет ряд преимуществ:

1. Использование focus score уменьшает вероятность появления ложноположительного результата за счет дополнительного настраивания сгенерированного окна на основе ошибок, полученных на первой стадии.
2. Соревновательная процедура обучения улучшает обобщающую способность модели, а также помогает модели быть более устойчивой.

Как было упомянуто ранее, каждый из декодировщиков независимо пытается восстановить входное окно временного ряда. Определим потерю восстановления (reconstruction loss) следующим образом:

$$l_1 = \|O_1 - W\|_2;$$

$$l_2 = \|O_2 - W\|_2.$$

Дополнительно необходимо ввести понятие соревновательной потери (adversarial loss). Во время второй фазы второй декодировщик старается отличить входное окно от окна, сгенерированного первым декодировщиком, путем максимизации  $\|\hat{O}_2 - W\|_2$  (соревновательная потеря). С другой стороны, первый декодировщик старается «обмануть» второй декодировщик путем генерации такого окна  $O_1$ , чтобы focus score был близок к нулю. То есть в процессе обучения нужно оптимизировать следующую величину:

$$\min_{\text{Decoder1}} \max_{\text{Decoder2}} \|\hat{O}_2 - W\|_2.$$

Таким образом, цель первого декодировщика — минимизировать потерю восстановления, в то время как целью второго является максимизация соревновательной потери.

$$l_1 = \|\hat{O}_2 - W\|_2;$$

$$l_2 = -\|\hat{O}_2 - W\|_2.$$

Тогда общая функция потерь будет иметь вид:

$$L_1 = \frac{1}{n} \|O_1 - W\|_2 + (1 - \frac{1}{n}) \|\hat{O}_2 - W\|_2;$$

$$L_2 = \frac{1}{n} \|O_2 - W\|_2 - (1 - \frac{1}{n}) \|\hat{O}_2 - W\|_2,$$

где  $n$  — номер эпохи. На начальных эпохах ошибке восстановления дается большой вес (за вес отвечает множитель  $\frac{1}{n}$ ). Это необходимо для того, чтобы focus score ко второй фазе получался небольшой, что, в свою очередь, повышает устойчивость модели.

### 3.4.3 Детектирование аномалий

Детектирование аномалий на тестовых данных осуществляется следующим образом. Сначала для тестовых данных  $(\hat{W}, \hat{C})$  считается anomaly score:

$$s = \frac{1}{2} \|O_1 - \hat{W}\|_2 + \frac{1}{2} \|\hat{O}_2 - \hat{W}\|_2.$$

На тесте данные рассматриваются только до текущего момента времени  $t$  и anomaly score  $s_i$  считается для каждого момента времени  $t$  и для каждой размерности временного ряда  $i$ . Текущая временная метка  $t$  помечается аномальной, если  $s_i$  превосходит некоторый порог, причем для каждой из размерностей  $i$  этот порог свой. Для подбора этих значений используется метод динамического подбора порогов POT (Peak Over Threshold) [Siffer et al., 2017]. Таким образом, наличие аномалии в момент времени  $t$  определяется следующим образом:

$$y = \vee_i y_i = \vee_i 1(s_i \geq POT(s_i)).$$

#### 3.4.4 Полученные результаты

В статье TranAD [Tuli et al., 2022] демонстрируются результаты, на основе которых данная модель лучше всех справляется с задачей детектирования аномалий на различных временных рядах. Однако особенность данных, рассматриваемых в статье, заключается в наличии разметки аномальных моментов, что позволяет явно замерить качество полученного решения.

Подробное описание данных было приведено выше в соответствующей секции. Для решения поставленной задачи в качестве обучающей и валидационной выборки использовались 9 многомерных временных рядов и 1 соответственно. Для тестирования модели были использованы 5 нейтральных временных рядов и 5 положительных.

Авторы статьи использовали в качестве временного ряда  $C$  окно фиксированного размера  $K = 10$ , а в качестве окна, которое подается на вход в window encoder — одну временную точку  $x_t \in \mathbb{R}^{62}$ . Для решения поставленной задачи размер окна был выбран равным 15.

Ошибка восстановления и соревновательная ошибка на обучающих и валидационных данных представлена на 6.

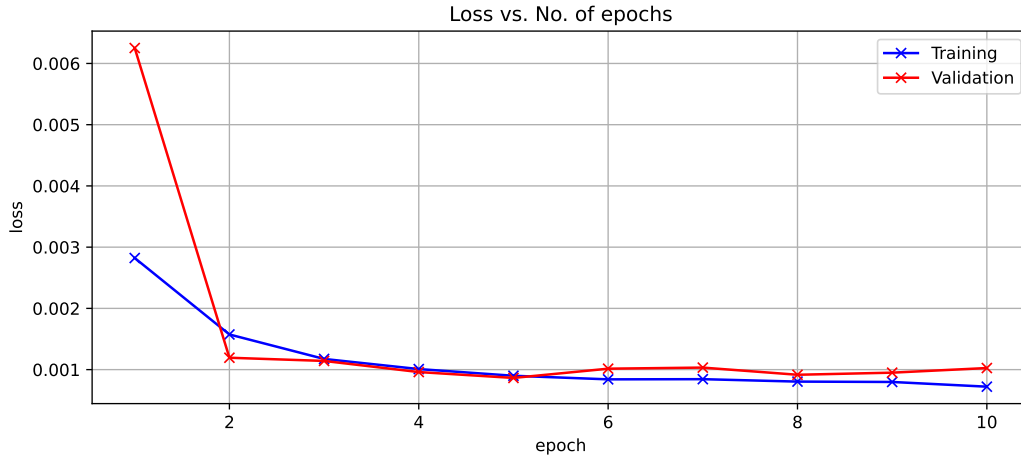


Рис. 6: Зависимость ошибки от номера эпохи на обучении

После обучения модели на 10 эпохах необходимо было подобрать пороги для каждой из размерностей с помощью алгоритма POT. У данного алгоритма есть гиперпараметр, который прямо коррелирует с тем, какой процент точек будет принят аномальными. Наша гипотеза заключается в том, что для нейтральных данных процент таких точек должен быть очень мал, исходя из чего данный параметр был подобран таким образом, чтобы на нейтральных тестовых данных доля аномальных точек не превосходила пяти процентов.

Для тестовых данных для каждой временной точки на основе порога можно получить ответ алгоритма — является ли данная точка аномальной или нет. Как было упомянуто ранее, разметка для сигналов ЭЭГ отсутствует, поэтому для оценки качества полученного решения были использованы косвенные метрики.

Таким образом, AUC ROC модели, получающей оценки принадлежности нейтральному классу описанным выше способом будет равен 0.8, что подтверждает эффективность предложенного решения. Общий процент аномальных точек в тестовых данных для нейтрального класса будет 5, а для положи-



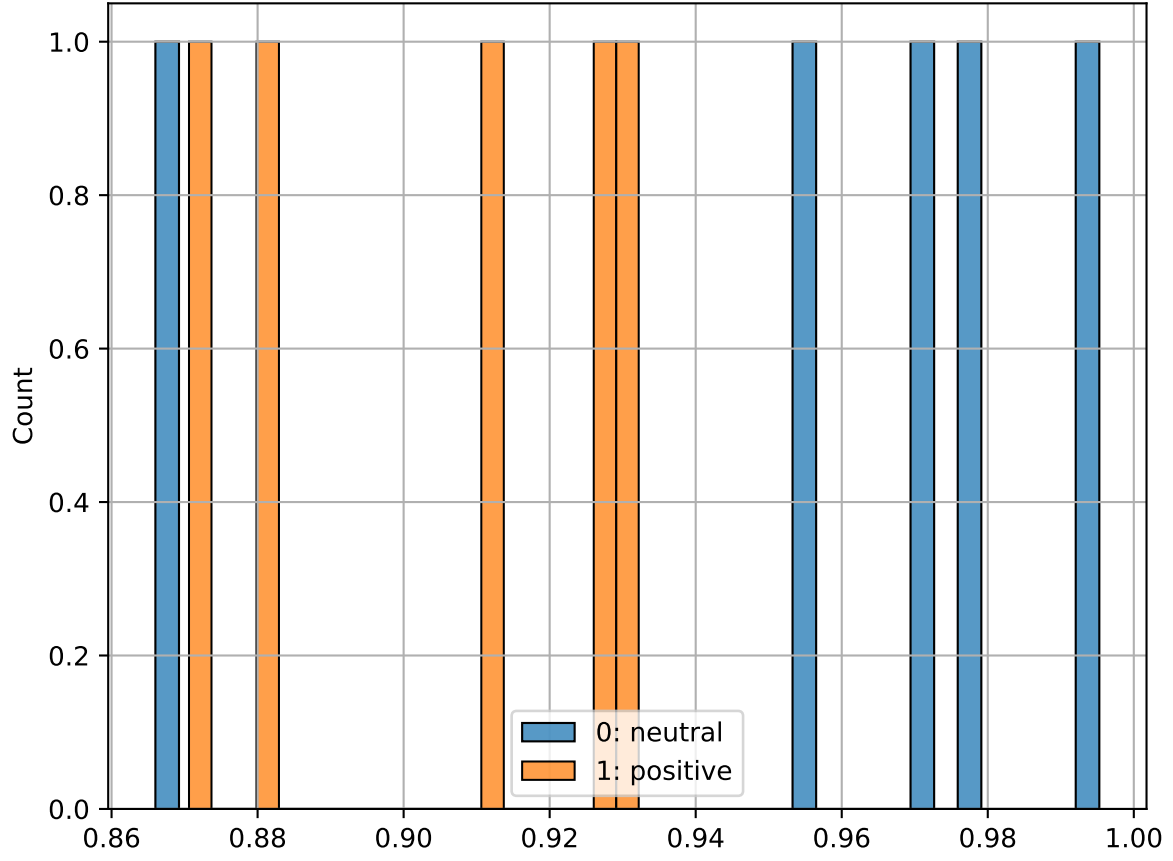


Рис. 7: Распределение процентов неаномальных точек для объектов тестовой выборки

тельного 10, что дополнительно демонстрирует, что модель научилась различать тенденции в рядах ЭЭГ, вызванных положительной сценой, от тенденций, вызванных нейтральной сценой.

### 3.5 Анализ выявленных аномалий

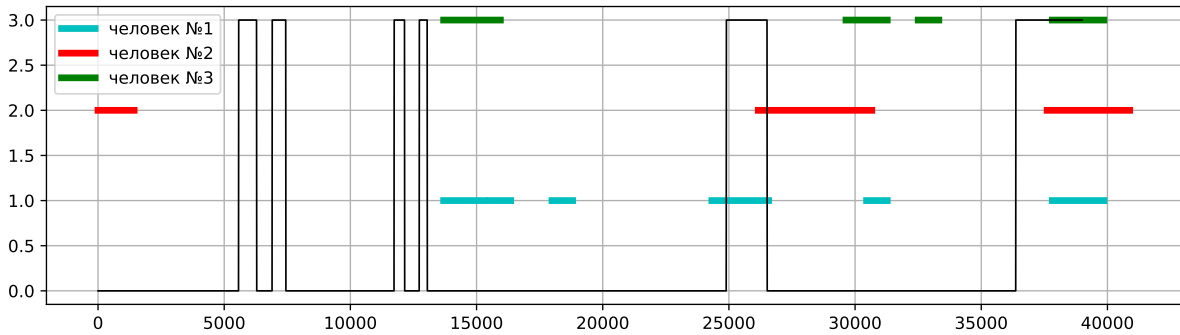


Рис. 8: Результаты независимых экспертов

Для того чтобы проинтерпретировать полученные результаты, рассмотрим реальный фрагмент положительной сцены из фильма "Затерянные в Таиланде" 2012 года, используемый в тестовых данных.

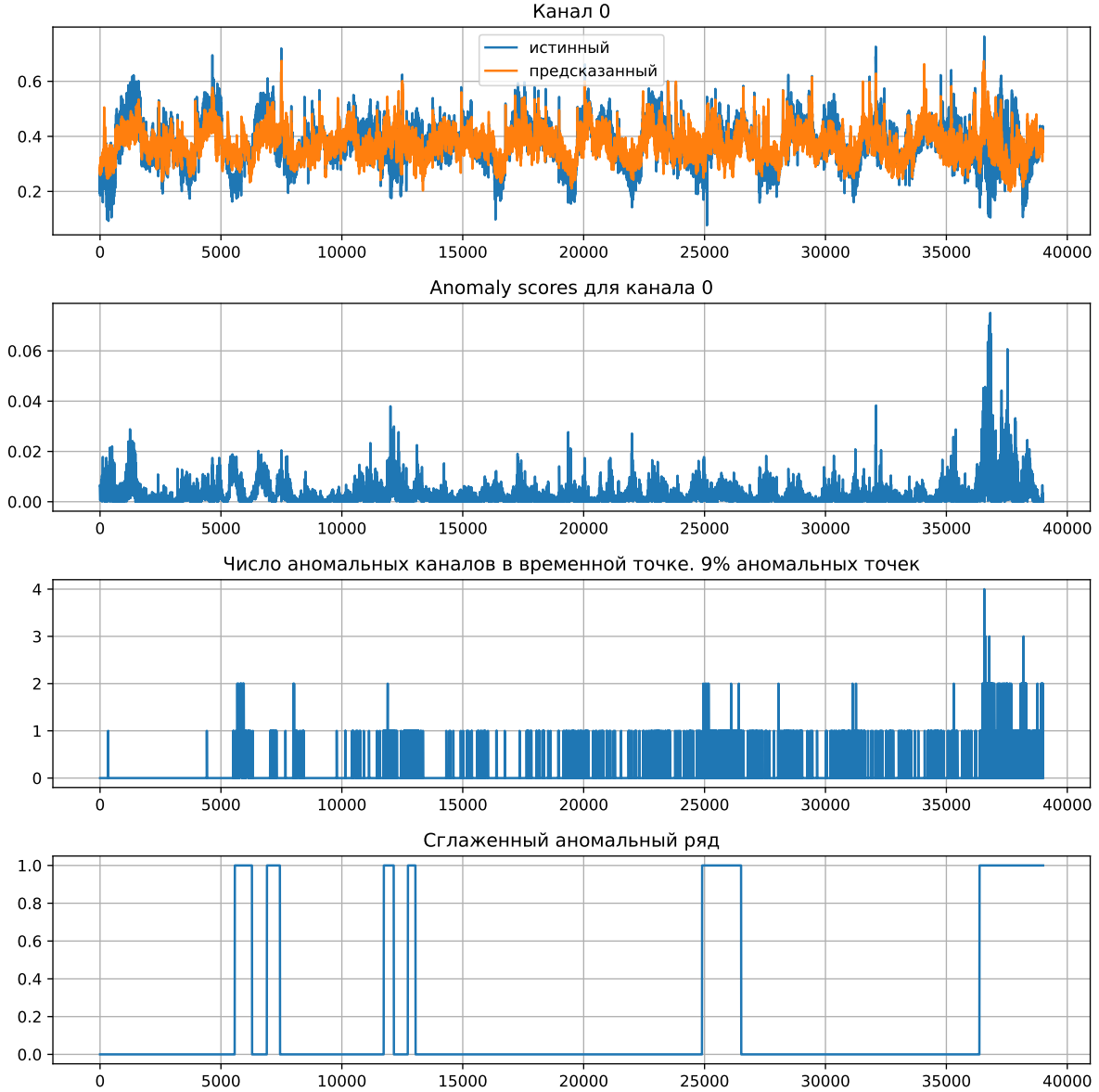


Рис. 9: Результаты работы алгоритма на положительном примере

Задача заключается в нахождении моментов, которые вызывают у зрителя самые яркие эмоции. Яркая эмоция обычно бывает непрерывной во времени, поэтому к массиву, в котором сохранены число аномальных каналов в данной временной точке, применяется сглаживание. Используется следующая эвристика: временную точку  $t$  будем считать аномальной, если в окрестности в 200 временных отсчетов по крайней мере 50 точек являются аномальными. Благодаря данному преобразованию был получен график, представленный на рисунке 9.

Для того чтобы определить адекватность полученных результатов трем независимым экспертам, не имеющим представлений о задаче, было предложено просмотреть данный фрагмент, переведенный на русский язык, и выделить отрезки, которые вызвали у них самые сильные эмоции. Полученные результаты представлены на рисунке 8.

Можно заметить, что все эксперты выделили практически один и тот же временной отрезок в конце сцены. Данный фрагмент соответствует самой смешной шутке в сцене. Модель также успешно справилась с этим отрезком и детектировала его.

## 4 Выводы

В данной работе были получены следующие результаты:

1. Исследованы различные методы детектирования аномалий, проанализированы сильные и слабые стороны каждого из подходов.
2. Подобраны и проанализированы данные, подходящие под решение поставленной задачи.
3. Была создана собственная реализация алгоритма, описанного в статье. Был применен новый подход к выявлению эмоций на основе детектирования аномалий, который продемонстрировал интерпретируемые результаты.

## Список литературы

- Aurobinda Routray-Monalisa Sarma Satyajit Nayak, Bingi Nagesh. A human-computer interaction framework for emotion recognition through time-series thermal video sequences. *Computers Electrical Engineering*, 93, 2021. doi:10.1016/j.compeleceng.2021.107280.
- Paolo M Russ Valentina Colonnello, Katia Mattarozzi. Emotion recognition in medical students: effects of facial appearance and care schema activation. *Medical Education*, 53:195–205, 2018. doi:10.1111/medu.13760.
- Xianglin Pan Longhui Qiu Xiang Feng, Yaojia Wei and Yongmei Ma. Academic emotion classification and recognition method for large-scale online learning environment—based on a-cnn and lstm-att deep learning pipeline method. *International Journal of Environmental Research and Public Health*, 2020. doi:10.3390/ijerph17061941.
- Zhi Yao Yuxin Ren Yuanhao Wu Qiqi Fan Erkang Fu, Xi Li. Personnel emotion recognition model for internet of vehicles security monitoring in community public space. *EURASIP Journal on Advances in Signal Processing*, Article number: 81, 2021. doi:10.21203/rs.3.rs-797632/v1.
- Xiao Sun; Yezhen Song; Meng Wang. Toward sensing emotions with deep visual analysis: A long-term psychological modeling approach. *IEEE MultiMedia*, 27:18–27, 2020. doi:10.1109/MMUL.2020.3025161.
- Kori M. Inkpen Regan L. Mandryk, M. Stella Atkins. A continuous and objective evaluation of emotional experience with interactive play environments. *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1027–1036, 2006. doi:10.1145/1124772.1124926.
- Tim Indersmitten Mahendra T. Bhati Dorothy Charbonnier Ruben C. Gur Bruce I. Turetsky, Christian G. Kohler. Facial emotion recognition in schizophrenia: when and why does it go awry? *Schizophrenia Research*, 94:253–263, 2007. doi:10.1016/j.schres.2007.05.001.
- M. Yang Z. Li Z. Li D. Liao X. Xu L. Shu, J. Xie and X. Yang. A review of emotion recognition using physiological signals. *Sensors*, 18:2074, 2018. doi:10.3390/s18072074.
- Xingguang Li Yujian Cai and Jinsong Li. Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. *Sensors*, 2023. doi:10.3390/s23052455.
- William Holderbaum Stephen Haben, Marcus Voß. Time series forecasting: Core concepts and definitions. *Core Concepts and Methods in Load Forecasting*, pages 55–66, 2023. doi:10.1007/978-3-031-27852-5\_5.
- B. B. Zaidan A. A. Zaidan Mohamed Hamada. A systematic review for human eeg brain signals based emotion classification, feature extraction, brain condition, group comparison. *Journal of Medical Systems*, 2018. doi:10.1007/s10916-018-1020-8.
- W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, pages 1–13, 2018. ISSN 2168-2267. doi:10.1109/TCYB.2018.2797176.
- Bernhard Schölkopf, Robert C. Williamson, Alex J. Smola, John Shawe-Taylor, and John C. Platt. Support vector method for novelty detection. *Advances in Neural Information Processing Systems 12 (NIPS 1999)*, pages 582–588, 1999. doi:10.1007/s10916-018-1020-8.
- Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008.
- Reda A. El-Khoribi Salma Alhagry, Aly Aly Fahmy. Emotion recognition based on eeg using lstm recurrent neural network. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 8, 2017. doi:10.14569/IJACSA.2017.081046.

- Lukas Dominique Josef Fiederer Martin Glasstetter Katharina Eggersperger Michael Tangermann Frank Hutter Wolfram Burgard Tonio Ball Robin Tibor Schirrmeister, Jost Tobias Springenberg. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38:5391–5420, 2017. doi:10.1002/hbm.23730.
- Sae-Byuk Shin Yea-Hoon Kwon and Shin-Dug Kim. Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (cnn) model for emotion recognition system. *Sensors*, page 1383, 2018. doi:10.3390/s18051383.
- R.K. Tripathy Manish Sharma U. Rajendra Acharya Daksh Maheshwari, S.K. Ghosh. Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel eeg signals. *Computers in Biology and Medicine*, 134, 2021. doi:10.1016/j.compbimed.2021.104428.
- Pachori R. B. Sharma, R. and P. Sircar. Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomedical Signal Processing and Control*, 58, 2020. doi:10.1016/j.bspc.2020.101867.
- Wei-Long Zheng Yun Luo, Si-Yang Zhang and Bao-Liang Lu. Wgan domain adaptation for eeg-based emotion recognition. *ICONIP 2018: Neural Information Processing*, pages 275—286, 2018. doi:10.1007/978-3-030-04221-9\_25.
- Shreshth Tuli, Giuliano Casale, and Nicholas R Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *Proceedings of VLDB*, 15(6):1201–1214, 2022. doi:10.48550/arXiv.2201.07284.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.
- Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, and Christine Largouët. Anomaly detection in streams with extreme value theory. pages 1067–1075, 2017. URL <http://dblp.uni-trier.de/db/conf/kdd/kdd2017.html#SifferFTL17>.